Underdetermined Source Separation using Mixtures of Warped Laplacians

Nikolaos Mitianoudis and Tania Stathaki

Communications and Signal Processing Group,
Imperial College London,
Exhibition Road, London SW7 2AZ, UK
n.mitianoudis@imperial.ac.uk

Abstract. In a previous work, the authors have introduced a Mixture of Laplacians model in order to cluster the observed data into the sound sources that exist in an underdetermined two-sensor setup. Since the assumed linear support of the ordinary Laplacian distribution is not valid to model angular quantities, such as the Direction of Arrival to the set of sensors, the authors investigate the performance of a Mixture of Warped Laplacians to perform efficient source separation with promising results.

1 Introduction

Assume that a set of M microphones $\boldsymbol{x}(n) = [x_1(n), \ldots, x_M(n)]^T$ observes a set of N sound sources $\boldsymbol{s}(n) = [s_1(n), \ldots, s_N(n)]^T$. The case of instantaneous mixing, i.e. each sensor captures a scaled version of each signal with no delay in transmission, will be considered with negligible additive noise. The instantaneous mixing model can thus be expressed in mathematical terms, as follows:

$$\boldsymbol{x}(n) = \boldsymbol{A}\boldsymbol{s}(n) \tag{1}$$

where $\boldsymbol{A}$ represents the *mixing matrix* and n the sample index. The blind source separation problem provides an estimate of the source signals $\boldsymbol{s}$, based on the observed microphone signals and some general source statistical profile.

The underdetermined source separation problem ($M < N$) is a challenging problem. In this case, the estimation of the mixing matrix A is not sufficient for the estimation of nonGaussian source signals $\boldsymbol{s}$, as the pseudo-inverse of $\boldsymbol{A}$ can not provide a valid solution. Hence, this blind estimation problem can be divided into two sub-problems: i) estimating the mixing matrix $\boldsymbol{A}$ and ii) estimating the source signals $\boldsymbol{s}$.

In this study, we will assume a two sensor instantaneous mixing approach. The combination of several instruments into a stereo mixture in a recording studio follows the instantaneous mixing model of (1). The proposed approach can thus be used to decompose a studio recording into the separate instruments that exist in the mixture for many possible applications, such as music transcription, object-based audio coding and audio remixing.

The solution of the two above problems is facilitated by moving to a sparser representation of the data, such as the *Modified Discrete Cosine Transform* (MDCT). In the case of sparse sources, the density of the data in the mixture space shows a tendency to cluster along the directions of the mixing matrix columns. It has been demonstrated [6] that the phase difference θ_n between the two sensors can be used to identify and separate the sources in the mixture.

$$\theta_n = \text{atan}\frac{x_2(n)}{x_1(n)} \tag{2}$$

Using the phase difference information between the two sensors is equivalent to mapping all the observed data points on the unit-circle. The strong super-Gaussian characteristics of the individual components in the MDCT domain are preserved in the angle representation θ_n. We can also define the amplitude r_n of each point $\boldsymbol{x}(n)$, as follows:

$$r_n = \sqrt{x_1(n)^2 + x_2(n)^2} \tag{3}$$

In a previous work [6], we proposed a clustering approach on the observed θ_n to perform source separation. In order to model the sparse characteristics of the source distributions, we introduced the following Mixture of Laplacians (MoL) that was trained using an Expectation-Maximisation (EM) algorithm on the observed angles θ_n of the input data.

$$p(\theta_n) = \sum_{i=1}^{N} \alpha_i \mathcal{L}(\theta, c_i, m_i) = \sum_{i=1}^{N} \alpha_i c_i e^{-2c_i|\theta_n - m_i|} \tag{4}$$

where N is the number of the Laplacians in the mixture, m_i defines the mean and $c_i \in \mathbf{R}^+$ controls the "width" of the distribution. Once the model is trained each of the Laplacians of the MoL should be centred on the Direction of Arrival (DOA) of the sources in the majority of the cases, i.e. the angles denoted by the columns of the mixing matrix. One can perform separation using optimal detection approaches for the individual trained Laplacians.

There is a shortcoming in the previous assumed model. The model in (4) assumes a linear support for θ_n, which is not valid as the actual support for θ_n wraps around $\pm 90^o$. The linear support is not a problem if the sources are well contained within $[-90^o, 90^o]$. To overcome this problem, we proposed a strategy in [6], where in each update we check whether any of the centres are closer to any of the boundaries ($\pm 90^o$). In this case, all the data points and the estimated centres m_i are rotated, so that the affected boundary (-90^o or 90^o) is mapped to the middle of the centres m_i that feature the greatest distance. This seemed to alleviate the problem in the majority of cases, however, it still serves as a heuristic solution.

To address this problem in a more eloquent manner, one can introduce wrapped *distributions* to provide a more complete solution. In the literature, there exist several "circular" distributions, such as the von Mises distribution (also known as the circular normal distribution). However, this definition is

rather difficult to optimise in an EM perspective. In this study, we examine the use of an approximate warped-Laplacian distribution to model the periodicity of 180^o that exists in atan$(\cdot)$ with encouraging results.

2 A Mixture of Warped Laplacians

The observed angles θ_n of the input data can be modelled, as a Laplacian wrapped around the interval $[-90^o, 90^o]$ using the following additive model:

$$\begin{aligned}\mathcal{L}_w(\theta, c, m) &= \frac{1}{2T-1}\sum_{t=-T}^{T} ce^{-2c|\theta-m-180t|} \\ &= \frac{1}{2T-1}\sum_{t=-T}^{T} \mathcal{L}(\theta-180t, c, m) \qquad \forall\ \theta_n \in [-90^o, 90^o] \end{aligned} \tag{5}$$

where $T \in \mathbf{Z}^+$ denotes the number of ordinary Laplacians participating in the wrapped version. The above expression models the wrapped Laplacian by an ordinary Laplacian and its periodic repetitions by 180^o. This is an extension of the wrapped Gaussian distribution proposed by Smaragdis and Boufounos [7] for the Laplacian case. The addition of the wrapping of the distribution aims at mirroring the wrapping of the observed angles at $\pm 90^o$, due to the atan$(\cdot)$ function. In general, the model should have $T \to \infty$ components, however, it seems that in practice a small range of values for T can successfully approximate the full warped probability density function.

In a similar fashion to Gaussian Mixture Models (GMM), one can introduce *Mixture of Warped Laplacians* (MoWL) in order to model a mixture of angular or circular sparse signals. A *Mixture of Warped Laplacians* can thus be defined, as follows:

$$p(\theta) = \sum_{i=1}^{N} \alpha_i \mathcal{L}_w(\theta, c_i, m_i) = \sum_{i=1}^{N} \alpha_i \frac{1}{2T-1}\sum_{t=-T}^{T} c_i e^{-2c_i|\theta-m_i-180t|} \tag{6}$$

where α_i, m_i, c_i represent the weight, mean and width of each Laplacian respectively and all weights should sum up to one, i.e. $\sum_{i=1}^{N} \alpha_i = 1$. The *Expectation-Maximization* (EM) algorithm has been proposed as a valid method to train a mixture model [1]. Consequently, the EM can be employed to train a MoWL over a training set. We derive the EM algorithm, based on Bilmes's analysis [1] for the estimation of a GMM. Bilmes estimates Maximum Likelihood mixture density parameters using the EM [1]. Assuming K training samples for θ_n and Mixture of Warped Laplacians densities (6), the log-likelihood of these training samples θ_n takes the following form:

$$I(\alpha_i, c_i, m_i) = \sum_{n=1}^{K} \log \sum_{i=1}^{N} \alpha_i \mathcal{L}_w(\theta_n, c_i, m_i) \tag{7}$$

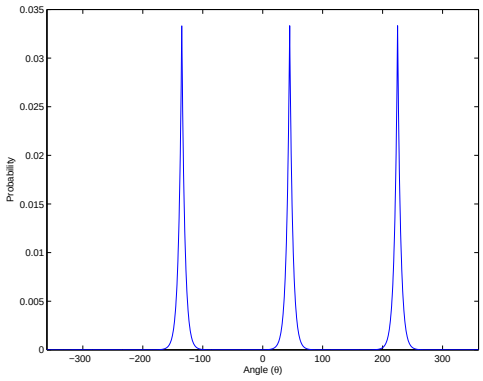

Fig. 1. An example of the Wrapped Laplacian for $T = [-1, 0, 1]$ $c = 0.01$ and $m = 45^o$.

Introducing unobserved data items that can identify the components that "generated" each data item, we can simplify the log-likelihood of (7) for Warped Laplacian Mixtures, as follows:

$$J(\alpha_i, c_i, m_i) = \sum_{n=1}^{K}\sum_{i=1}^{N}\left(\log\alpha_i - \log(2T+1) + \log\sum_{t=-T}^{T}\mathcal{L}(\theta - 180t, c_i, m_i)\right)p(i|\theta_n) \tag{8}$$

where $p(i|\theta_n)$ represents the probability of sample θ_n belonging to the i^{th} Laplacian of the MoWL. In a similar manner, we can also introduce unobserved data items to identify the individual Laplacian of the i^{th} Warped Laplacian that depends on θ_n.

$$H(\alpha_i, c_i, m_i) = \sum_{n=1}^{K}\sum_{i=1}^{N}(\log\alpha_i - \log(2T+1) + \log c_i \tag{9}$$

$$- \sum_{t=-T}^{T} 2c_i|\theta - 180t - m_i|p(t|i,\theta_n))p(i|\theta_n) \tag{10}$$

where $p(t|i,\theta_n)$ represents the probability of sample θ_n belonging to the i^{th} Warped Laplacian and the t^{th} individual Laplacian. The updates for $p(t|i,\theta_n)$, $p(i|\theta_n)$ and α_i can be given by the following equations:

$$p(t|i,\theta_n) = \frac{\mathcal{L}(\theta_n - t\pi, m_i, c_i)}{\sum_{t=-T}^{T}\mathcal{L}(\theta_n - 180t, m_i, c_i)} \tag{11}$$

$$p(i|\theta_n) = \frac{\alpha_i\mathcal{L}_w(\theta_n, m_i, c_i)}{\sum_{i=1}^{N}\alpha_i\mathcal{L}_w(\theta_n, m_i, c_i)} \tag{12}$$

$$\alpha_i \leftarrow \frac{1}{K}\sum_{n=1}^{K}p(i|\theta_n) \tag{13}$$

In a similar manner to [6], one can set $\partial H(\alpha_i, c_i, m_i)/\partial m_i = 0$ and solve for m_i for the recursive update for m_i, as follows:

$$\frac{\partial H}{\partial m_i} = \sum_{n=1}^{K} \sum_{t=-T}^{T} 2c_i \text{sgn}(\theta_n - 180t - m_i) p(t|i, \theta_n) p(i|\theta_n) = 0 \Rightarrow \tag{14}$$

$$\sum_{n=1}^{K} \sum_{t=-T}^{T} \frac{\theta_n - 180t}{|\theta_n - 180t - m_i|} p(t|i, \theta_n) p(i|\theta_n) = m_i \sum_{n=1}^{K} \sum_{t=-T}^{T} \frac{p(t|i, \theta_n) p(i|\theta_n)}{|\theta_n - 180t - m_i|} \Rightarrow \tag{15}$$

$$m_i \leftarrow \frac{\sum_{n=1}^{K} \sum_{t=-T}^{T} \frac{\theta_n - 180t}{|\theta_n - 180t - m_i|} p(t|i, \theta_n) p(i|\theta_n)}{\sum_{n=1}^{K} \sum_{t=-T}^{T} \frac{1}{|\theta_n - 180t - m_i|} p(t|i, \theta_n) p(i|\theta_n)} \tag{16}$$

Similarly, one can set $\partial H(\alpha_i, c_i, m_i)/\partial c_i = 0$, to solve for the estimate of c_i :

$$\frac{\partial H}{\partial c_i} = \sum_{n=1}^{K} (c_i^{-1} - 2 \sum_{t=-T}^{T} |\theta_n - 180t - m_i| p(t|i, \theta_n)) p(i|\theta_n) = 0 \Rightarrow \tag{17}$$

$$c_i \leftarrow \frac{\sum_{n=1}^{K} p(i|\theta_n)}{2 \sum_{n=1}^{K} \sum_{t=-T}^{T} |\theta_n - 180t - m_i| p(t|i, \theta_n) p(i|\theta_n)} \tag{18}$$

Once the MoWL is trained, optimal detection theory and the estimated individual Laplacians can be employed to provide estimates of the sources. The centre of each warped Laplacian m_i should represent a column of the mixing matrix A in the form of $[\cos(m_i) \ \sin(m_i)]^T$. Each warped Laplacian should model the statistics of each source in the transform domain and can be used to perform underdetermined source separation.

A "Winner takes all" strategy attributes each point (r_n, θ_n) to only one of the sources. This is performed by setting a hard threshold at the intersections between the trained Warped Laplacians. Consequently, the source separation problem becomes an *optimal decision* problem. The decision thresholds θ_{ij}^{opt} between the i-th and the j-th neighbouring Laplacians are given by the following equation:

$$\theta_{ij}^{opt} = \frac{\ln \frac{\alpha_i c_i}{\alpha_j c_j} + 2(c_i m_i + c_j m_j)}{2(c_i + c_j)} \tag{19}$$

Using these thresholds, the algorithm can attribute the points with $\theta_{ij}^{opt} < \theta_n < \theta_{jk}^{opt}$ to source j, where i, j, k are neighbouring Laplacians (sources). Having attributed the points $\boldsymbol{x}(n)$ to the N sources, using the proposed thresholding technique, the next step is to reconstruct the sources. Let $S_i \cap K$ represent the point indices that have been attributed to the i^{th} source. We initialise $u_i(n) = 0, \forall$ $n = 1, \ldots, K$ and $i = 1, \ldots, N$. The source reconstruction is performed by substituting:

$$u_i(S_i) = [\cos(m_i) \ \sin(m_i)] \boldsymbol{x}(S_i) \qquad \forall \ i = 1, \ldots, N \tag{20}$$

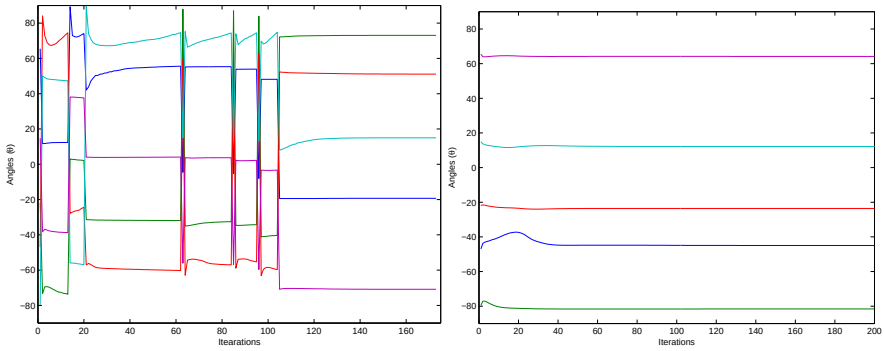

Fig. 2. Estimation of the mean using MoL with the shifting strategy (left) and the warped MoL (right).

3 Experiments

In this section, we evaluate the algorithm proposed in the previous section. We will use Hyvärinen's clustering approach [4], O'Grady and Pearlmutter's [5] Soft LOST algorithm's and the MoL-EM_Hard as proposed in a previous work [6], to demonstrate several trends using artificial mixtures or publicly available datasets[1]. In order to quantify the performance of the algorithms, we are estimating the *Signal-to-Distortion Ratio* (SDR) from the BSS_EVAL Toolbox [2]. The frame length for the MDCT analysis is set to 64 msec for the test signals sampled at 16 KHz and to 46.4 msec for those at 44.1 KHz. We initialise the parameters of the MoL and MoWL, as follows: $\alpha_i = 1/N$ and $c_i = 0.001$ and $T = [-1, 0, 1]$ (for MoWL only). The centres m_i were initialised in both cases using a *K-means* step. The initialisation of m_i is important, as if we choose two initial values for m_i that are really close, then it is very probable that the individual Laplacians may not converge to different clusters. To provide a more accurate estimation of m_i, training is initially performed using a "reduced" dataset, containing all points that satisfy $r_n > 0.2$, provided that the input signals are scaled to $[-1, 1]$. The second phase is to use the "complete" dataset to update the values for α_i and c_i.

3.1 Artificial Experiment

In this experiment, we use 5 solo audio uncorrelated recordings (a saxophone, an accordion, an acoustic guitar, a violin and a female voice) of sampling frequency 16 KHz and duration 8.19 msec. The mixing matrix is constructed as in (21), choosing the angles in Table 1. Two of the sources are placed close to the wrapping edges $(-80^o, 60^o)$ and three of them are placed rather closely at $-40^o, -20^o, 10^o$, in order to test the algorithm's resilience to the wrapping at

[1] All the experimental audio results are available online at:
`http://www.commsp.ee.ic.ac.uk/~nikolao/lmm.htm`

$\pm 90^o$. In Table 1, we can see the estimated angles of the original MoL_Hard with the shifting solution and the MoWL. In both cases, the algorithms estimate approximately the same means m_i, which are very close to the original ones. In Fig. 2, the convergence of the means m_i in the two cases is depicted. The proposed warped solution seems to converge smoothly and faster without the perturbations caused by the shifting solution in the previous algorithm. Note that Fig. 2(a) depicts the angles after the rotating steps to demonstrate the shifting of ψ_i in the original MoL solution. Their performance in terms of SDR is depicted in Table 2. Hyvärinen's approach is very prone to initialisation, however, the results are acquired using the best run of the algorithm. This could be equally avoided by using a K-means initialisation step. The Soft_Lost algorithm managed to separate the sources in most cases, however, there were some audible artifacts and clicks that reduced the calculated quality measure. To appreciate the results of this rather difficult problem, we can spot the improvement performed by the methods compared to the input signals. It seems that the proposed algorithm performs similarly to MoL_Hard and the Hyvärinen's approach, which implies that the proposed solution to approximate the wrapping of the pdf is valid.

$$A = \begin{bmatrix} \cos(\psi_1) & \cos(\psi_2) & \dots & \cos(\psi_N) \\ \sin(\psi_1) & \sin(\psi_2) & \dots & \sin(\psi_N) \end{bmatrix} \tag{21}$$

Table 1. The five angles used in the artificial experiment and their estimates using the MoL and MoWL approaches.

	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5
`Original`	-80^o	-40^o	-20^o	10^o	60^o
`Estimated MoL`	-81.52^o	-45.45^o	-23.45^o	12.59^o	64.18^o
`Estimated MoWL`	-81.59^o	-44.98^o	-23.59^o	12.18^o	64.19^o

3.2 Real Recording

In this section, we tested the algorithms with the *Groove* dataset, available by (BASS-dB) [3], sampled at 44.1 KHz. The "Groove" dataset features four widely spaced sources: bass (far left), distortion guitar (center left), clean guitar (center right) and drums (far right). In Table 2, we can see the results for the four methods in terms of SDR. The proposed MoWL approach managed to perform similarly to the previous MoL_EM, despite the small spacing of the sources and the source being placed at the edges of the solution space, which implies that the warped Laplacian model manages to model the warping of θ_n without any additional steps. The proposed MoL approaches managed to outperform Hyvärinen and Soft_LOST approach for the majority of the sources. Again, the LOST approach still introduces several audio artifacts and clicks.

Table 2. The proposed MoWL approach is compared in terms of SDR (dB) with MoL-EM_hard, Hyvärinen's, soft_LOST and the average SDR of the mixtures.

	Artificial experiment					Groove Dataset			
	s_1	s_2	s_3	s_4	s_5	s_1	s_2	s_3	s_4
Mixed Signals	-6.00	-13.37	-26.26	-6.67	-6.81	-30.02	-10.25	-6.14	-21.24
MoWL-EM_hard	6.07	-2.11	5.62	4.09	6.15	4.32	-4.35	-1.16	3.27
MoL-EM_hard	6.69	0.32	7.66	3.65	6.03	2.85	-4.47	-0.86	3.28
Hyvärinen	6.53	-1.16	7.60	4.14	5.79	3.79	-3.72	-1.13	1.49
soft_LOST	4.58	-4.01	5.09	1.67	3.93	4.54	-5.77	-1.74	3.62

4 Conclusions

The problem of underdetermined source separation is examined in this study. In a previous work, we proposed to address the two-sensor problem by clustering using a Mixture of Laplacian approach on the source Direction of Arrival (DOA) θ_n to the sensors. In this study, we address the problem of wrapping of θ_n using a Warped Mixture of Laplacians approach. The new proposed approach features similar performance and faster convergence to MoL_hard and seems to be able to separate sources that are close to the boundaries ($\pm 90^0$) without any extra trick and therefore serves as a valid solution to the problem.

References

1. J.A. Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian Mixture and Hidden Mixture Models," Tech. Rep., Dep. of Electrical Eng. and Computer Science, U.C. Berkeley, California, 1998.
2. C. Févotte, R. Gribonval, and E. Vincent, "BSS EVAL Toolbox User Guide," Tech. Rep., IRISA Technical Report 1706, Rennes, France, April 2005, http://www.irisa.fr/metiss/bss eval/.
3. E. Vincent, R. Gribonval, C. Fevotte, A. Nesbit, M.D. Plumbley, M.E. Davies, and L. Daudet, "BASS-dB: the blind audio source separation evaluation database," Available at `http://bass-db.gforge.inria.fr/BASS-dB/`.
4. A. Hyvärinen, "Independent component analysis in the presence of Gaussian noise by maximizing joint likelihood," *Neurocomputing*, vol. 22, pp. 49–67, 1998.
5. P.D. O'Grady and B.A. Pearlmutter, "Soft-LOST: EM on a mixture of oriented lines," in *Proc. Int. Conf. on Independent Component Analysis and Blind Source Separation*, Granada, Spain, 2004, pp. 428–435.
6. N. Mitianoudis and T. Stathaki, "Underdetermined Source Separation using Laplacian Mixture Models", *IEEE Trans. on Audio, Speech and Language*, to appear.
7. P. Smaragdis and P. Boufounos, "Position and Trajectory Learning for Microphone Arrays", *IEEE Trans. Audio, Speech and Language Processing*, vol. 15, no. 1, pp. 358 - 368, 2007.